The Annals of Statistics
1987, Vol. 15, No. 2, 626–641

ONE-STEP *L*-ESTIMATORS FOR THE LINEAR MODEL[1]

BY A. H. WELSH

University of Chicago

We propose and investigate the asymptotic properties of a class of estimators of the regression parameter in the general linear model. The estimators depend on a preliminary estimate of the regression parameter and the residuals based on it. For the location model, the estimators are linear combinations of the order statistics and the robustness and efficiency properties of this class of estimators carry over to the general linear model. The estimators settle the doubts raised by Bickel (1973) about the feasibility of the construction of a general class of reparametrization invariant estimators of a regression parameter which are linear combinations of order statistics in the location problem.

1. Introduction. Suppose that we observe $Y_1, Y_2, \dots, Y_n$, where

$$Y_j = x_j'\theta_0 + e_j, \qquad 1 \le j \le n, \tag{1.1}$$

with $\{x_j' = (1, x_{j2}, \dots, x_{jp})\}$ a sequence of known p-vectors ($p > 1$), $\theta_0 \in \mathscr{R}^p$ an unknown parameter to be estimated and $\{e_j\}$ a sequence of independent and identically distributed random variables with common distribution function F. In general, if F is not symmetric, an intercept is not identifiable so without loss of generality we put $\theta_{01} = 0$ and absorb the intercept into the errors $\{e_j\}$. The regressors may depend on n but we suppress this dependence for notational simplicity. When $p = 1$, (1.1) corresponds to the location problem. A rich class of location estimators can be represented or approximated by linear functions of the order statistics $Y_{n1} \le Y_{n2} \le \cdots \le Y_{nn}$ otherwise known as L-estimators. The purpose of this paper is to generalize the class of L-estimators to the linear model (1.1).

Following Serfling (1980, page 265), we adopt a von Mises functional representation as the definition of L-estimators of location. For any distribution function G, let $G^{-1}(t) = \inf\{s\colon G(s) \ge t\}$ and define a functional

$$T(G) = \int_0^1 G^{-1}(u)\, dH(u) + \sum_{i=1}^{m} w_i G^{-1}(q_i),$$

where $H(u) = \int_0^u h(t)\, dt$, $0 < u < 1$, is the distribution function of a finite signed measure on $(0,1)$, $w_1, \dots, w_m$ are constant weights and $0 < q_1 < \cdots < q_m < 1$ for $m < \infty$. We call h the weight function and without loss of generality suppose that $\int_0^1 dH(u) + \sum_{i=1}^m w_i = 1$. Adopting the convention that in the absence of explicit limits of integration the range of integration is to be taken as $-\infty$ to ∞,

---

Received February 1986; revised August 1986.

[1] Research supported in part by National Science Foundation grant DMS-8601732.

AMS 1980 *subject classifications.* 62G05, 60F05, 62J05.

Key words and phrases. Efficient estimation, linear combinations of order statistics, linear model, quantiles, robust estimation.

we have

$$T(G) = \int y\, dH(G(y)) + \sum_{i=1}^{m} w_i G^{-1}(q_i). \tag{1.2}$$

Let F_n denote the empirical distribution function of $e_1, \ldots, e_n$ (which in the location problem is equivalent to that of $Y_1, \ldots, Y_n$). Then the class of L-estimators of location is defined by

$$T(F_n) \doteq n^{-1} \sum_{j=1}^{n} h(j/n) Y_{nj} + \sum_{i=1}^{m} w_i Y_{nl(q_i)}, \tag{1.3}$$

where

$$l(q) = \begin{cases} nq, & \text{if } nq \text{ is an integer,} \\ [nq] + 1, & \text{otherwise, } 0 \le q \le 1. \end{cases}$$

We refer to the first term in (1.3) as the smooth term and the second term as the quantile term; if h vanishes near zero and one the resulting L-estimator is trimmed, otherwise it is untrimmed.

For the linear model (1.1), Bickel (1973) constructed an interesting class of one-step L-estimators by examining the component quantile functions. While the estimators enjoy the requisite asymptotic properties, their calculation is complex (involving p orderings) and they are not invariant to reparametrization. Koenker and Bassett (1978) introduced regression quantiles and used them to develop analogues of the quantiles, systematic statistics such as the trimean and the trimmed mean. Other one-step analogues of the trimmed mean have been discussed by Ruppert and Carroll (1980) and Welsh (1987). The present paper extends the approach of Welsh (1987) to construct analogues of the general class of L-estimators given by (1.2). The resulting estimators provide one-step versions of the quantiles and systematic statistics, include the trimmed regression parameter estimator of Welsh (1987) as a special case and also include such estimators as the smoothly trimmed means such as that with

$$h(u) = 6(u - \alpha)(\beta - u) I(\alpha \le u \le \beta)$$

or that proposed by Stigler (1973), the redescending L-estimators for which h is negative near α and β as in the weight function

$$h(u) = \{\sin 4\pi g(u) / \tan \pi g(u)\} I(\alpha \le u \le \beta),$$

where $g(u) = \{(u - a)/(\beta - \alpha)\} - \frac{1}{2}$, and the asymptotically efficient L-estimators for F known with density f for which $h(u) = \phi(u)\phi''(u) / \int_0^1 \phi''(t)\, dt$, where $\phi(u) = f(F^{-1}(u))$, $0 \le u \le 1$, provided

$$0 < \int_0^1 \phi''(t)\, dt < \infty.$$

These estimators require only a single ordering for their calculation and provided the initial estimator is regression and scale equivariant and invariant to reparametrization so is the resulting L-estimator, thus settling the doubts raised by Bickel (1973) about the feasibility of such a construction.

As in Welsh (1987), the construction involves the influence curve. Boos (1979) and Serfling (1980, page 279) give conditions under which the distribution of

$n^{1/2}(T(F_n) - T(F))$ is asymptotically equivalent to that of $n^{-1/2}\Sigma_{j=1}^n \psi(e_j)$, where

$$\psi(z) = -\int \{I(z \le y) - F(y)\} h(F(y))\, dy - \sum_{i=1}^m \{w_i/\phi(q_i)\}\{I(z \le F^{-1}(q_i)) - q_i\}$$

is the influence curve of $T(\cdot)$ at F. Except for an obvious matrix normalization, we construct an estimator which when appropriately centered is asymptotically equivalent to $n^{-1/2}\Sigma_{j=1}^n x_j \psi(e_j)$ by considering $n^{-1/2}\Sigma_{j=1}^n x_j \psi_n(r_j)$, where $r_j = Y_j - x_j'\theta_n$, $1 \le j \le n$, are the residuals from a preliminary estimator θ_n and ψ_n is an estimator of ψ, and then correcting for the centering and the use of residuals involving θ_n. More formally, if $\Sigma_{j=1}^n x_{jk} = 0$, $k = 2, \dots, p$, we construct an estimator λ_n such that

$$\lambda_n = \theta_0 + \begin{pmatrix} T(F) \\ 0 \end{pmatrix} + \left(\sum_{j=1}^n x_j x_j'\right)^{-1} \sum_{j=1}^n x_j \psi(e_j) + o_p(n^{-1/2})$$

holds by (essentially) replacing θ_0 by θ_n and $T(F)$, ψ and $e_1, \dots, e_n$ by their empirical counterparts based on the residuals $r_1, \dots, r_n$. In fact, it is convenient to also replace $\Sigma_{j=1}^n x_j x_j'$ by a different but asymptotically equivalent matrix. Specifically, if G_n is the empirical distribution function of $r_1, \dots, r_n$, we are led to

$$\begin{aligned} \lambda_n &= \theta_n + \begin{pmatrix} T(G_n) \\ 0 \end{pmatrix} - C_n^- \sum_{j=1}^n x_j \Bigg[\int \{I(r_j \le y) - G_n(y)\} h(G_n(y))\, dy \\ &\qquad + \sum_{i=1}^m \{w_i/\phi_n(q_i)\}\{I(r_j \le G_n^{-1}(q_i)) - q_i\} \Bigg] \\ (1.4) \quad &= C_n^- \sum_{j=1}^n x_j \Bigg[\int y\, dH(G_n(y)) + x_j'\theta_n h(G_n(r_j)) \\ &\qquad - \int \{I(r_j \le y) - G_n(y)\} h(G_n(y))\, dy \Bigg] \\ &\qquad + \sum_{i=1}^m w_i C_n^- \sum_{j=1}^n x_j \left[G_n^{-1}(q_i) + x_j'\theta_n - \{I(r_j \le G_n^{-1}(q_i)) - q_i\}/\phi_n(q_i) \right], \end{aligned}$$

where C_n^- is a generalized inverse of $C_n = \Sigma_{j=1}^n x_j x_j'\{h(G_n(r_j)) + \Sigma_{i=1}^m w_i\}$ and ϕ_n is any pointwise consistent estimator of ϕ. We discuss possible choices of θ_n and ϕ_n in Section 2. Notice that $\lambda_{n1} = T(G_n)$ if we put $G_n(G_n^{-1}(q)) = q$, $0 < q < 1$, and $\theta_{n1} = 0$. Our main result (Theorem 1) in Section 2 establishes that λ_n is a generalization of $T(F_n)$ to the linear model problem and incidentally also gives conditions under which C_n is asymptotically nonsingular.

Insight into the nature of λ_n may be gained by letting either $\Sigma_{i=1}^m w_i = 0$ or $h(u) = 0$ and considering the smooth and quantile terms separately. Suppose that $h(u) = 0$ for $u < \alpha$ or $u > \beta$, $0 \le \alpha < \beta \le 1$, and h is differentiable at all but a finite number of points so we can integrate the smooth term in the influence curve by parts. Then if we adopt the convention that $G_n(G_n^{-1}(q)) = q$,

$0 \le q \le 1$, for n large enough the smooth term in (1.4) can be written as

$$\tau_n = \theta_n + C_n^- \sum_{j=1}^n x_j \Bigg[h(\alpha) G_n^{-1}(\alpha)\{I(r_j \le G_n^{-1}(\alpha)) - \alpha\} + r_j h(G_n(r_j))$$
$$+ \sum_{k=1}^n r_k \{I(r_j \le r_k) - G_n(r_k)\} h'(G_n(r_k))$$
$$+ h(\beta) G_n^{-1}(\beta)\{I(r_j > G_n^{-1}(\beta)) - (1-\beta)\}\Bigg],$$

where the first term vanishes if $\alpha = 0$ and the last term vanishes if $\beta = 1$. We refer to the first and fourth terms as edge-effect terms. If we define a sequence $\{D(1), \dots, D(n)\}$ by

$$r_{nj} = r_{D(j)}, \qquad 1 \le j \le n,$$

where $r_{n1} \le r_{n2} \le \cdots \le r_{nn}$, we can write the smooth term as

$$\tau_n = \theta_n + C_n^- \left(\sum_{j=1}^{l(\alpha)} x_{D(j)} - \alpha \sum_{j=1}^n x_j \right) r_{nl(\alpha)} h(\alpha) + C_n^- \sum_{j=l(\alpha)+1}^{l(\beta)} x_{D(j)} r_{D(j)} h(j/n)$$
$$+ C_n^- \sum_{k=l(\alpha)+1}^{l(\beta)} \left(\sum_{j=1}^k x_{D(j)} - (k/n) \sum_{j=1}^n x_j \right) r_{nk} h'(k/n)$$
$$+ C_n^- \left(\sum_{j=l(\beta)+1}^n x_{D(j)} - (1-\beta) \sum_{j=1}^n x_j \right) r_{nl(\beta)} h(\beta),$$

where the relevant edge-effect terms vanish if either $\alpha = 0$ or $\beta = 1$, and the quantile term as

$$\theta_n + \sum_{i=1}^m w_i \left[C_n^- \sum_{j=1}^n x_j r_{nl(q_i)} - C_n^- \left(\sum_{j=1}^{l(q_i)} x_{D(j)} - q_i \sum_{j=1}^n x_j \right) \Big/ \phi_n(q_i) \right].$$

Thus, the one-step estimator of a single regression quantile, $0 < q < 1$, is just

$$\lambda_n = \theta_n + \begin{pmatrix} G_n^{-1}(q) \\ 0 \end{pmatrix} - C_n^- \sum_{j=1}^n x_j \{I(r_j \le G_n^{-1}(q)) - q\} \Big/ \phi_n(q),$$

which involves a shift of the intercept and a method-of-scoring step toward a solution of the regression quantile "normal" equations [Koenker and Bassett (1978)]. The unnormalized weight function $h(u) = I(\alpha \le u \le \beta)$ recovers the trimmed mean of Welsh (1987) and the comments of Welsh on the number of observations trimmed by this estimator apply to the other trimmed L-estimators λ_n. The weight functions of the smoothly trimmed means and the redescending estimators satisfy $h(\alpha) = h(\beta) = 0$ so that for these estimators the edge-effect terms always vanish; however, the term in h' does not always vanish for these estimators though it does for the trimmed mean. Finally, in the location problem, the term involving h' in the integrated form of the estimator vanishes and we have

$$\begin{aligned}\lambda_n = {} & n^{-1}(l(\alpha) - n\alpha)Y_{nl(\alpha)}h(\alpha) \\ & + n^{-1} \sum_{j=l(\alpha)+1}^{l(\beta)} Y_{nj}h(j/n) + n^{-1}(n\beta - l(\beta))Y_{nl(\beta)}h(\beta) \\ & + \sum_{i=1}^{m} w_i\Big[Y_{nl(q_i)} - n^{-1}(l(q_i) - nq_i)/\phi_n(q_i)\Big],\end{aligned}$$

which reduces to (1.3) if nq_i is an integer, $1 \le i \le m$, and either λ_n is untrimmed or λ_n is trimmed and in addition either $n\alpha$ and $n\beta$ are integers or $h(\alpha) = h(\beta) = 0$. It is sensible to take $\theta_{n1} = 0$ and just neglect the edge-effect terms to preserve the relation $\lambda_{n1} = T(G_n)$.

The results of this paper are presented and discussed in the next section and proved in Section 3. All probability statements are made at the true parameter value θ_0 and all limits are taken as $n \to \infty$.

2. Results. The asymptotic theory of location L-estimators has been investigated by among others Chernoff, Gastwirth and Johns (1967), Shorack (1969, 1972), Stigler (1969, 1974), Boos (1979) and Serfling (1980, page 271); an interesting discussion of these approaches is given by Serfling (1980, page 271). Our results and the methods we use to prove them are extensions of those of Boos (1979) and Serfling (1980, page 284).

We impose throughout the following basic conditions on the model (1.1):

(i) $n^{1/2}(\theta_n - \theta_0)$ is bounded in probability;

(ii) $x_{j1} = 1$ for all j, $\sum_{j=1}^n x_{jk} = 0$, $k = 2, \dots, p$, for each n, and there exists a positive definite matrix Γ such that

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} x_j x_j' = \Gamma;$$

(iii) the density f is uniformly continuous, positive and bounded;

and

(iv) $\phi_n(q_i) - \phi(q_i) \to_P 0$, $1 \le i \le m$.

The first two conditions of (ii) ensure that the slope is identifiable even when F is asymmetric; the second of these conditions entails no loss of generality because we may simply replace each x_j by $(1, x_{j1} - \bar{x}_1, \dots, x_{jp} - \bar{x}_p)'$, where $\bar{x}_k = n^{-1}\sum_{j=1}^n x_{jk}$, $2 \le k \le p$, $1 \le j \le n$. Since $\theta_{01} = 0$, it is sensible to put $\theta_{n1} = 0$ (after calculating θ_n with an unknown intercept) so that condition (i) is just a condition on the preliminary slope estimator. Condition (iii) is stronger than the conditions required on F for location L-estimators but is required for the weak convergence of empirical processes based on regression residuals. We discuss conditions (i) and (iv) further at the end of this section but note that (iv) will often be unnecessary and for particular choices of ϕ_n may be implied by the first three conditions. As in the location problem, the theoretical treatment of trimmed L-estimators is different from that of untrimmed L-estimators with different conditions required on F and h in each case. For trimmed L-estimators, we impose

(A) the weight function h is bounded, continuous a.e. Lebesgue and satisfies $h(u) = 0$ for $u < \alpha$ or $u > \beta$, where $0 < \alpha < \beta < 1$;

while for untrimmed L-estimators we impose

(B)(i) $\int [F(x+\varepsilon)\{1 - F(x-\varepsilon)\}]^{1/2}\, dx < \infty$ for some $\varepsilon > 0$

and

(B)(ii) the weight function h is bounded and continuous.

Serfling (1980, page 284) does not require h to be bounded in (B)(ii) but this modification seems minor [see Stigler (1974, page 677)] and the remaining conditions (A) and (B) are essentially those imposed by Boos (1979) and Serfling (1980, page 271) with, of course, the modification that (i)–(iv) hold. For practical purposes [see Stigler (1974, page 686)] condition (B)(i) involves essentially $\operatorname{Var}(e_1) < \infty$. We denote by (B′) the same set of conditions with

$$\operatorname{Var}(e_1) < \infty.$$

We will require (B′) to consistently estimate the asymptotic variance. Bickel (1973) requires $E|e| < \infty$ and $E\psi(e)^2 < \infty$ for the untrimmed case and whenever h puts positive weight on the extremes, the latter condition will require $\operatorname{Var}(e_1) < \infty$. Condition (B)(ii) is weaker than the corresponding condition in Bickel (1973).

The first theorem establishes that λ_n is indeed a generalization of $T(F_n)$. The proof which uses ideas from Boos (1979), Serfling (1980, page 279) and Bickel (1973) is given in Section 3.

THEOREM 1. *Suppose that the basic conditions* (i)–(iv) *hold and that either* (A) *or* (B) *holds. Then with* $\lambda_0 = \theta_0 + (T(F), 0, \dots, 0)' \in \mathscr{R}^p$, *we have*

$$n^{1/2}(\lambda_n - \lambda_0) - n^{1/2}\Gamma^{-1}\sum_{j=1}^{n} x_j\psi(e_j) \to_P 0.$$

It follows that provided $E\psi(e)^2 < \infty$,

$$n^{1/2}(\lambda_n - \lambda_0) \to_D N\big(0, E\psi(e)^2\Gamma^{-1}\big).$$

If F is symmetric, an intercept is identifiable and $T(F)$ being the center of symmetry coincides with the intercept. For the trimmed mean, condition (iii) of Theorem 1 can be weakened slightly [see Welsh (1987)].

A natural estimator of the quantity $E\psi(e)^2$ may be derived by considering $n^{-1}\sum_{j=1}^{n}\psi_n(r_j)^2$. Writing $G_n(G_n^{-1}(q)) = q$, $0 < q < 1$, we obtain

$$\begin{aligned} v_n = {} & \int\int \{G_n(z \wedge y) - G_n(z)G_n(y)\}h(G_n(z))h(G_n(y))\, dy\, dz \\ & + 2\sum_{i=1}^{m}\{w_i/\phi_n(q_i)\}\int\{q_i \wedge G_n(y) - q_iG_n(y)\}h(G_n(y))\, dy \\ & + \sum_{i=1}^{m}\sum_{k=1}^{m}\{w_iw_k/\phi_n(q_i)\phi_n(q_k)\}(q_i \wedge q_k - q_iq_k), \end{aligned}$$

where $z \wedge y = \min(z, y)$. If h' exists at all but a finite number of points, we can integrate the smooth term in the influence curve by parts and for n large enough using again the convention that $G_n(G_n^{-1}(q)) = q, 0 < q < 1$, we can conveniently write the first two terms of v_n as

$$\begin{aligned} n^{-1} \sum_{j=1}^{n} \left(r_{nj}h(j/n) - \bar{r}\right)^2 &+ \alpha(1-\alpha)r_{nl(\alpha)}^2 h(\alpha)^2 + \beta(1-\beta)r_{nl(\beta)}^2 h(\beta)^2 \\ &-2\alpha(1-\beta)r_{nl(\alpha)}r_{nl(\beta)}h(\alpha)h(\beta) - 2\alpha r_{nl(\alpha)}\bar{r}h(\alpha) - 2(1-\beta)r_{nl(\beta)}\bar{r}h(\beta) \\ &+n^{-2}\sum_{j=1}^{n}\sum_{k=1}^{n} r_{nj}r_{nk}n^{-2}(n(j \wedge k) - jk)h'(j/n)h'(k/n) \\ &+2n^{-1}\sum_{j=1}^{n} r_{nj}h'(j/n)n^{-1}\left(\sum_{k=1}^{j} r_{nk}h(k/n) - j\bar{r}\right) \\ &+2r_{nl(\alpha)}h(\alpha)n^{-1}\sum_{j=1}^{n} r_{nj}(\alpha \wedge j/n - \alpha j/n)h'(j/n) \\ &-2r_{nl(\beta)}h(\beta)n^{-1}\sum_{j=1}^{n} r_{nj}(\beta \wedge j/n - \beta j/n)h'(j/n) \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^{m} \{w_i/\phi_n(q_i)\}\Bigg[n^{-1}\sum_{j=1}^{l(q_i)} r_{nj}h(j/n) &- q_i\bar{r} + (\alpha \wedge q_i - \alpha q_i)r_{nl(\alpha)}h(\alpha) \\ &-(\beta \wedge q_i - \beta q_i)r_{nl(\beta)}h(\beta) \\ &+n^{-1}\sum_{j=1}^{n} r_{nj}(q_i \wedge j/n - q_i j/n)h'(j/n)\Bigg], \end{aligned}$$

respectively, where $\bar{r} = n^{-1}\Sigma_{j=1}^{n} r_{nj}h(j/n)$, and $h(u) = 0$ for $u < \alpha$ or $u > \beta$, $0 \le \alpha < \beta \le 1$. With $h(u) = I(\alpha \le u \le \beta)$, $0 < \alpha < \beta < 1$, the estimator is equivalent though not identical to $(\beta - \alpha)^{-2}$ times the variance estimator for the trimmed mean in Welsh (1987).

Theorem 2 below gives conditions under which v_n is consistent for $E\psi(e)^2$. A similar estimator has been investigated in the location problem by Gardiner and Sen (1979).

THEOREM 2. *Suppose that the conditions* (i)–(iv) *hold and that either* (A) *or* (B′) *holds. Then*

$$v_n\left(n^{-1}\sum_{j=1}^{n} x_j x_j'\right)^{-1} \to_P E\psi(e)^2\Gamma^{-1}.$$

In the location problem, conditions (i) and (ii) can be omitted and (iii) can be weakened.

One-step estimators are sensitive (particularly in small samples) to the initial estimator so that the choice of initial estimator requires some care. If conditions (ii) and (B′) hold, the least-squares estimator satisfies condition (i) so that it may be used as an initial estimator for the untrimmed L-estimators when asymptotic efficiency is sought. However, from the robustness viewpoint, a more robust preliminary estimator such as the least-absolute deviations estimator [which, moreover, does not require $\operatorname{Var}(e_1) < \infty$ to hold in order to satisfy (i); see Corollary 4 of Ruppert and Carroll (1980)] should be used. Other M-estimators may be used as initial estimators [see Carroll (1979)] but in general the fact that a concommitant scale estimate is required makes these estimators unattractive in the present context. While the use of a robust initial estimator is desirable for the trimmed L-estimators, a prudent approach would be to use such an initial estimator for the untrimmed L-estimators too.

One method of obtaining a consistent estimator of ϕ is to define

$$\phi_{ns}(q) = n^{-1/2}s\{G_n^{-1}(q + n^{-1/2}s(1-q)) - G_n^{-1}(q - n^{-1/2}sq)\}^{-1},$$
$$0 < s < n^{1/2}, 0 < q < 1.$$

Notice that

$$\begin{aligned}\phi_{ns}(q)^{-1} - \phi(q)^{-1} &= (n^{1/2}/s)G_n^{-1}(q + n^{-1/2}s(1-q)) - (1-q)\phi(q)^{-1}\\ &\quad -(n^{1/2}/s)G_n^{-1}(q - n^{-1/2}sq) - q\phi(q)^{-1},\end{aligned}$$

so that if conditions (i)–(iii) hold and $s_n \to_P s_0$ for some positive constant s_0, we have that for fixed $0 < q < 1$, $0 < \varepsilon \le q - n^{-1/2}s_n q \le 1 - \varepsilon < 1$ for n large enough so

$$\begin{aligned}&(n^{1/2}/s_n)\big|G_n^{-1}(q - n^{-1/2}s_n q) + n^{-1/2}s_n q\phi(q)^{-1}\\ &\qquad -\{F_n(F^{-1}(q)) - q\}/\phi(q) - F^{-1}(q) + \bar{x}'(\theta_n - \theta_0)\big|\\ &\quad\le (n^{1/2}/s_n)\big|G_n^{-1}(q - n^{-1/2}s_n q) - F^{-1}(q - n^{-1/2}s_n q)\\ &\qquad\quad +\{F_n(F^{-1}(q - n^{-1/2}s_n q)) - (q - n^{-1/2}s_n q)\}/\\ &\qquad\qquad \phi(q - n^{-1/2}s_{nq}) + \bar{x}'(\theta_n - \theta_0)\big|\\ &\qquad +(n^{1/2}/s_n)\big|\{F_n(F^{-1}(q - n^{-1/2}s_n q)) - (q - n^{-1/2}s_n q)\}/\\ &\qquad\qquad \phi(q - n^{-1/2}s_n q) - \{F_n(F^{-1}(q)) - q\}/\phi(q)\big|\\ &\qquad +\big|(n^{1/2}/s_n)\{F^{-1}(q - n^{-1/2}s_n q) - F^{-1}(q)\} + q\phi(q)^{-1}\big|\\ &\quad\to_P 0,\end{aligned}$$

by Lemma 4.6 of Bickel (1973), the weak convergence properties of the empirical process and the one-term Taylor series expansion for F^{-1}. Similarly,

$$\begin{aligned}&(n^{1/2}/s_n)\big|G_n^{-1}(q + n^{-1/2}s_n(1-q)) - n^{1/2}s_n(1-q)\phi(q)^{-1}\\ &\qquad -\{F_n(F^{-1}(q)) - q\}/\phi(q) - F^{-1}(q) + \bar{x}'(\theta_n - \theta_0)\big| \to_P 0,\end{aligned}$$

whence

$$|\phi_{ns_n}(q) - \phi(q)| \to_P 0.$$

Now s is effectively a window-width so the usual problems of selecting a suitable s arise. Note, however, that a reasonable rather than an "optimal" (in some sense) choice is required and that the choice may be determined after examining the data, perhaps even fairly subjectively. The above approach is closely related to ideas in Siddiqui (1960) and Geertsema (1970). It is clear that approaches based on the kernel and other methods of density estimation may also be useful; see for example Falk (1986). Finally, we note that the problem of estimating ϕ can be avoided altogether if we restrict attention to the rich class of L-estimators defined by the smooth term alone and avoid the use of the one step analogues of the systematic statistics and the Winsorized mean.

3. Proofs. We begin by proving two preliminary lemmas which will be needed for the proof of Theorem 1 under condition (B).

LEMMA 1. *Let* $\{c_{jn}\}$ *be any sequence of constants such that* $\lim_{n\to\infty} n^{-1}\sum_{j=1}^{n} c_{jn}^2 < \infty$ *and* $\lim_{n\to\infty}\max_{1\le j\le n} n^{-1/2}|c_{jn}| = 0$. *Then if* (ii), (iii) *and* (B) *hold*,

$$\sup_{|t|\le n^{-1/2}M} n^{-1/2}\int\left|\sum_{j=1}^{n} c_{jn}\{I(e_j\le y+x_j't)-F(y+x_j't)\}\right|dy$$

is bounded in probability for any $M<\infty$.

PROOF. Write $c_j = c_{jn}$, $1\le j\le n$, for convenience and, writing $c_j = c_j^+ - c_j^-$ if necessary, without loss of generality take $c_j\ge 0$, $1\le j\le n$. The result will follow if we can show that for each fixed t_0 and $\delta>0$,

$$T_1 = n^{-1/2}\int\left|\sum_{j=1}^{n} c_j\{I(e_j\le y+n^{-1/2}x_j't_0)-F(y+n^{-1/2}x_j't_0)\}\right|dy$$

and

$$\begin{aligned}\sup_{|t-t_0|\le\delta} n^{-1/2}\int\bigg|\sum_{j=1}^{n} c_j\{&I(e_j\le y+n^{-1/2}x_j't)-I(e_j\le y+n^{-1/2}x_j't_0)\\&+F(y+n^{-1/2}x_j't_0)-F(y+n^{-1/2}x_j't)\}\bigg|dy\\
\le{}& n^{-1/2}\int\left|\sum_{j=1}^{n} c_j\{I(e_j\le y+n^{-1/2}x_j't_0+\delta_j)-F(y+n^{-1/2}x_j't_0+\delta_j)\}\right|dy\\
&+n^{-1/2}\int\left|\sum_{j=1}^{n} c_j\{I(e_j\le y+n^{-1/2}x_j't_0-\delta_j)+F(y+n^{-1/2}x_j't_0-\delta_j\}\right|dy\\
&+2n^{-1/2}\int\left|\sum_{j=1}^{n} c_j\{F(y+n^{-1/2}x_j't_0+\delta_j)-F(y+n^{-1/2}x_j't_0-\delta_j)\}\right|dy\\
={}&T_{21}+T_{22}+T_{23},\end{aligned}$$

where $\delta_j = n^{-1/2}|x_j|\delta$, $1\le j\le n$, are bounded in probability.

Now, arguing as in Lemma 8.2.4D of Serfling (1980),

$$ET_1 \le \left(n^{-1}\sum_{j=1}^{n} c_j^2\right)^{1/2} \max_{1\le j\le n} \int \Big[F\big(y + n^{-1/2}x_j't_0\big) \times \Big\{1 - F\big(y + n^{-1/2}x_j't_0\big)\Big\}\Big]^{1/2} dy,$$

which is bounded so that T_1 is bounded in probability. Similarly, T_{21} and T_{22} are bounded in probability. Next, note that

$$\begin{aligned} &n^{-1/2}\int \sum_{j=1}^{n} c_j\Big\{F\big(y + n^{-1/2}x_j't_0 + \delta_j\big) - F(y)\Big\}\,dy \\ &\quad = n^{-1}\sum_{j=1}^{n} c_j\big(x_j't_0 + |x_j|\delta\big)\int \left\{\frac{F\big(y + n^{-1/2}x_j't_0 + \delta_j\big) - F(y)}{n^{-1/2}x_j't_0 + \delta_j} - f(y)\right\} dy \\ &\qquad + n^{-1}\sum_{j=1}^{n} c_j\big(x_j't_0 + |x_j|\delta\big), \end{aligned}$$

which is bounded so that T_{23} is bounded and the lemma obtains. □

LEMMA 2. *Under the conditions of Lemma* 1,

$$\sup_{|t|\le n^{-1/2}M} n^{-1/2}\left|\sum_{j=1}^{n} c_{jn}\int \big\{I(e_j \le y + x_j't) - F(y + x_j't) - I(e_j \le y) + F(y)\big\} h(F(y))\,dy\right| \to_P 0,$$

for any $M < \infty$.

PROOF. As before, let $c_j = c_{jn} \ge 0$, $1 \le j \le n$. By standard arguments (which are similar to those used in the proof of Lemma 1), it is enough to show that for each fixed t_0,

$$T_1 = n^{-1/2}\sum_{j=1}^{n} c_j \int \Big\{I\big(e_j \le y + n^{-1/2}x_j't_0\big) - F\big(y + n^{-1/2}x_j't_0\big) - I(e_j \le y) + F(y)\Big\} h(F(y))\,dy \to_P 0$$

and that for each $\delta > 0$,

$$T_2 = n^{-1/2}\left|\sum_{j=1}^{n} c_j \int \Big\{F\big(y + n^{-1/2}x_j't_0 + \delta_j\big) - F\big(y + n^{-1/2}x_j't_0 - \delta_j\big)\Big\} h(F(y))\,dy\right| = \delta O(1),$$

where $\delta_j = n^{-1/2}|x_j|\delta$, $1 \le j \le n$. However,

$$ET_1^2 \le n^{-1} \sum_{j=1}^{n} c_j^2 E\left[\int \left\{I\left(e_j \le y + n^{-1/2}x_j' t_0\right) - I(e_j \le y)\right\} h(F(y))\, dy\right]^2$$

$$\le Kn^{-1} \sum_{j=1}^{n} c_j^2 \int \left|F\left(y + n^{-1/2}x_j' t_0\right) - F(y)\right| dy \to 0,$$

by the dominated convergence theorem, so that $T_1 \to_P 0$. Also, a similar argument to that used to handle T_{23} in the proof of Lemma 1 yields the desired result for T_2 and hence the lemma. □

PROOF OF THEOREM 1. The result will follow if we can show that both

$$n^{-1/2}C_n(\lambda_n - \lambda_0) - n^{-1/2} \sum_{j=1}^{n} x_j \psi(e_j) \to_P 0 \quad \text{and} \quad n^{-1}C_n - \Gamma \to_P 0$$

hold. To this end, it is enough to show that for a fixed q, $0 < q < 1$,

$$\begin{aligned} n^{-1/2} \sum_{j=1}^{n} x_j\Big[&G_n^{-1}(q) - F^{-1}(q) + x_j'(\theta_n - \theta_0) \\ &- \left\{I\left(r_j \le G_n^{-1}(q)\right) - q\right\}/\phi_n(q) \\ &- \left\{I\left(e_j \le F^{-1}(q)\right) - q\right\}/\phi(q)\Big] \to_P 0, \end{aligned} \tag{3.1}$$

$$\begin{aligned} n^{-1/2} \sum_{j=1}^{n} x_j\Big[&\int y\, d\{H(G_n(y)) - H(F(y))\} + x_j'(\theta_n - \theta_0)h\left(G_n(r_j)\right) \\ &- \int \{I(r_j \le y) - G_n(y)\} h(G_n(y))\, dy \\ &+ \int \{I(e_j \le y) - F(y)\} h(F(y))\, dy\Big] \to_P 0, \end{aligned} \tag{3.2}$$

and

$$n^{-1} \sum_{j=1}^{n} x_j x_j' \left\{h\left(G_n(r_j)\right) - \int h(F(y))\, dF(y)\right\} \to_P 0. \tag{3.3}$$

We may write (3.1) as

$$\begin{aligned} -n^{-1/2} \sum_{j=1}^{n} x_j\Big[&I\left(r_j \le G_n^{-1}(q)\right) - I\left(e_j \le F^{-1}(q)\right) \\ &-\phi(q)\left\{G_n^{-1}(q) - F^{-1}(q) + x_j'(\theta_n - \theta_0)\right\}\Big]/\phi(q) \\ + \left\{\phi(q)^{-1} - \phi_n(q)^{-1}\right\}&n^{-1/2} \sum_{j=1}^{n} x_j\left\{I\left(r_j \le G_n^{-1}(q)\right) - q\right\}. \end{aligned}$$

The first term converges in probability to zero by a result of Koul (1969) and Bickel (1973) [which is stated as Lemma 1 in Welsh (1987)]. It follows from the same result that $n^{-1/2}\sum_{j=1}^{n} x_j\{I(r_j \le G_n^{-1}(q)) - q\}$ is bounded in probability and hence that (3.1) obtains.

It is convenient to prove that (3.2) holds componentwise. Let d_j denote any fixed component of x_j, $1 \le j \le n$; writing $d_j = d_j^+ - d_j^-$ if necessary, it is clear that there is no loss of generality involved in assuming $d_j \ge 0$, $1 \le j \le n$. For

$\bar{d}_n = n^{-1}\Sigma_{j=1}^n d_j > 0$, put

$$Q_n(y) = (n\bar{d}_n)^{-1} \sum_{j=1}^n d_j I(r_j \le y), \qquad P_n(y) = (n\bar{d}_n)^{-1} \sum_{j=1}^n d_j I(e_j \le y)$$

and

$$\bar{P}_n(y) = (n\bar{d}_n)^{-1} \sum_{j=1}^n d_j F\big(y + x_j'(\theta_n - \theta_0)\big)$$

and note that if $d_j = 1$, $1 \le j \le n$, $Q_n = G_n$, $P_n = F_n$ and $\bar{P}_n = \bar{F}_n$. The Koul–Bickel result referred to above establishes that

$$\sup_y n^{1/2}|Q_n(y) - \bar{P}_n(y) - P_n(y) + F(y)| \to_P 0. \tag{3.4}$$

In this notation, (3.2) will hold if we can show that

$$\begin{aligned}
R_1 &= n^{1/2}\Big[\int y\, d\{H(G_n(y)) - H(F(y))\} \\
&\qquad + \int \{\bar{F}_n(y) + F_n(y) - 2F(y)\} h(G_n(y))\, dy\Big] \\
&= -n^{1/2}\int \big[H(G_n(y)) - H(F(y)) - \{\bar{F}_n(y) + F_n(y) - 2F(y)\}h(G_n(y))\big]\, dy \\
&= -n^{1/2}\int W_{G_n,F}(y)\{G_n(y) - F(y)\}\, dy \\
&\qquad -n^{1/2}\int \{G_n(y) - \bar{F}_n(y) - F_n(y) + F(y)\}h(F(y))\, dy \\
&\qquad +n^{1/2}\int \{\bar{F}_n(y) + F_n(y) - 2F(y)\}\{h(G_n(y)) - h(F(y))\}\, dy \\
&= R_{11} + R_{12} + R_{13} \to_P 0,
\end{aligned}$$

where

$$W_{G,F}(y) = \begin{cases} \{H(G(y)) - H(F(y))\}/\{G(y) - F(y)\} - h(F(y)), \\ \qquad \text{if } G(y) \ne F(y), \\ 0, \qquad \text{otherwise,} \end{cases}$$

$$\begin{aligned}
R_2 &= n^{-1/2} \sum_{j=1}^n d_j x_j'(\theta_n - \theta_0) h\big(G_n(r_j)\big) - \bar{d}_n n^{1/2} \int \{\bar{P}_n(y) - F(y)\} h(F(y))\, dy \\
&= n^{1/2}(\theta_n - \theta_0)' n^{-1} \sum_{j=1}^n d_j x_j \Big[h\big(G_n(r_j)\big) - \int h(F(y))\, dF(y)\Big] \\
&\qquad -n^{-1/2} \sum_{j=1}^n d_j \int \{F\big(y + x_j'(\theta_n - \theta_0)\big) \\
&\qquad\qquad -F(y) - x_j'(\theta_n - \theta_0) f(y)\} h(F(y))\, dy \\
&= R_{21} + R_{22} \to_P 0,
\end{aligned}$$

and

$$R_3 = n^{1/2}\int\{\bar{P}_n(y) + P_n(y) - 2F(y)\}h(F(y))\,dy$$
$$-n^{1/2}\int\{Q_n(y) - G_n(y) + \bar{F}_n(y) + F_n(y) - 2F(y)\}h(G_n(y))\,dy$$
$$= n^{1/2}\int\{\bar{P}_n(y) + P_n(y) - 2F(y)\}\{h(F(y)) - h(G_n(y))\}\,dy$$
$$-n^{1/2}\int\{Q_n(y) - \bar{P}_n(y) - P_n(y) + F(y)\}h(G_n(y))\,dy$$
$$+n^{1/2}\int\{G_n(y) - \bar{F}_n(y) - F_n(y) + F(y)\}h(G_n(y))\,dy$$
$$= R_{31} + R_{32} + R_{33} \to_P 0.$$

Suppose first that (A) holds. Then as in Serfling (1980, page 281) there exist $-\infty < a < b < \infty$ such that for $\sup_y|G(y) - F(y)| < \min\{\alpha, 1-\beta\}$,

$$W_{G,F}(y) = h(G(y)) = h(F(y)) = 0$$

for $y < a$ or $y > b$. That is, for n large enough, we can truncate the limits of integration in many of the terms above and then apply the dominated convergence theorem. Specifically,

$$|R_{11}| \le \sup_y n^{1/2}|G_n(y) - F(y)|\int_a^b\left|W_{G_n,F}(y)\right|dy \to_P 0,$$

by the Koul–Bickel result (3.4) and the dominated convergence theorem; for details see Serfling (1980, page 281). A similar argument applies to R_{13} and R_{31}. Next, letting K denote a generic positive constant,

$$|R_{12}| \le K(b-a)\sup_y n^{1/2}\left|G_n(y) - \bar{F}_n(y) - F_n(y) + F(y)\right| \to_P 0,$$

by (3.4) and a similar argument applies to R_{22}, R_{32} and R_{33}. To show that the remaining term $R_{21} \to_P 0$, it is enough to prove that

$$\begin{aligned}\int h(G_n(y))\,dQ_n(y) - \int h(F(y))\,dF(y)\\ = \int_0^1\left\{h\left(G_n\left(Q_n^{-1}(u)\right)\right) - h(u)\right\}du \to_P 0,\end{aligned} \tag{3.5}$$

where to preserve notational simplicity, we omit explicit reference to the fact that the weights $\{d_j\}$ used in the construction of Q_n here are different from those used before. It is not hard to show using (3.4) and arguments similar to those involved in Proposition 4.1 and Lemma 4.6 of Bickel (1973) that

$$G_n\left(Q_n^{-1}(u)\right) - u \to_P 0, \qquad 0 < u < 1,$$

and then to apply the dominated convergence theorem to obtain $R_{21} \to_P 0$.

Now suppose that (B) holds. Then

$$|R_{11}| \le \sup_y\left|W_{G_n,F}(y)\right|n^{1/2}\int|G_n(y) - F(y)|\,dy \to_P 0,$$

by the Koul–Bickel result, the continuity of h and Lemma 1. Similarly, both R_{13} and R_{31} converge in probability to zero. It follows from Lemma 2 that $R_{12} \to_P 0$ and from Lemma 1, the Koul–Bickel result and Lemma 2 that both $R_{32} \to_P 0$ and $R_{33} \to_P 0$. Clearly, $R_{22} \to_P 0$ as in the proof of Lemma 1 and $R_{21} \to_P 0$ as before.

Finally, we can express (3.3) in the form of (3.5) and apply the arguments used above to complete the proof of the theorem. □

PROOF OF THEOREM 2. It is enough to show that $v_n \to_P E\psi(e)^2$ and this in turn will hold if we can show that

$$\begin{aligned}(3.6)\qquad &\int\int \{G_n(z \wedge y) - G_n(y)G_n(z)\}h(G_n(z))h(G_n(y))\,dy\,dz \\ &\qquad \to_P \int\int \{F(z \wedge y) - F(y)F(z)\}h(F(y))h(F(z))\,dy\,dz\end{aligned}$$

and that, for each fixed q, $0 < q < 1$,

$$\int \{q \wedge G_n(y) - qG_n(y)\}h(G_n(y))\,dy \to_P \int \{q \wedge F(y) - qF(y)\}h(F(y))\,dy.$$

Suppose first that (A) holds. Then arguing as in the proof of Theorem 1, there exists $-\infty < a < b < \infty$ such that for n sufficiently large, we can restrict the range of integration of all the above integrals to (a, b). Then the desired result is obtained from the Koul–Bickel result and the dominated convergence theorem.

Now suppose that (B′) holds. Then it follows from the Koul–Bickel result that for each y and z,

$$\begin{aligned}&\{G_n(z \wedge y) - G_n(z)G_n(y)\}h(G_n(z))h(G_n(y)) \\ &\qquad \to_P \{F(z \wedge y) - F(z)F(y)\}h(F(z))h(F(y)).\end{aligned}$$

Moreover,

$$|\{G_n(z \wedge y) - G_n(z)G_n(y)\}h(G_n(z))h(G_n(y))| \le G_n(z \wedge y) - G_n(z)G_n(y)$$

and clearly for each y and z,

$$G_n(z \wedge y) - G_n(z)G_n(y) \to_P F(z \wedge y) - F(z)F(y)$$

and

$$\begin{aligned}\int\int \{G_n(z \wedge y) - G_n(y)G_n(z)\}\,dy\,dz &= n^{-1}\sum_{j=1}^{n}\left(r_j - n^{-1}\sum_{k=1}^{n} r_k\right)^2 \\ &\to_P \operatorname{Var}(e_1) \\ &= \int\int \{F(z \wedge y) - F(y)F(z)\}\,dy\,dz,\end{aligned}$$

by a result of Hoeffding [Proposition 5 of Stigler (1974)], so that (3.6) follows from the extended dominated convergence theorem [Royden (1968, page 89)].

Finally, we can apply a similar argument to (3.7) since

$$\begin{aligned}
&\int \{q \wedge G_n(y) - qG_n(y)\}\, dy \\
&= q\int_{G_n^{-1}(q)}^{\infty} \{1 - G_n(y)\}\, dy + (1-q)\int_{-\infty}^{G_n^{-1}(y)} G_n(y)\, dy \\
&= G_n^{-1}(q)\{G_n(G_n^{-1}(q)) - q\} + \int_{G_n^{-1}(q)}^{\infty} y\, dG_n(y) \\
&\to_P \int_{F^{-1}(q)}^{\infty} dF(y) \\
&= \int \{q \wedge F(y) - qF(y)\}\, dy,
\end{aligned}$$

by the Koul–Bickel result, Lemma 4.6 of Bickel (1973) and by a slightly modified version of the argument of Lemma A.4 of Ruppert and Carroll (1980).

The result obtains. □

Acknowledgments. I am grateful to Stephen M. Stigler for helpful comments and to an Associate Editor and a referee for constructive suggestions. Jana Jurečková and Roger Koenker brought additional references to my attention.

REFERENCES

BICKEL, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597–616.

BOOS, D. D. (1979). A differential for *L*-statistics. *Ann. Statist.* **7** 955–959.

CARROLL, R. J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *J. Amer. Statist. Assoc.* **74** 674–679.

CHERNOFF, H., GASTWIRTH, J. L. and JOHNS, M. V., JR. (1967). Asymptotic distribution of linear combinations of order statistics, with applications to estimation. *Ann. Math. Statist.* **38** 52–72.

FALK, M. (1986). On the estimation of the quantile density function. *Statist. Probab. Lett.* **4** 69–73.

GARDINER, J. C. and SEN, P. K. (1979). Asymptotic normality of a variance estimator of a linear combination of a function of order statistics. *Z. Wahrsch. verw. Gebiete* **50** 205–221.

GEERTSEMA, J. C. (1970). Sequential confidence intervals based on rank tests. *Ann. Math. Statist.* **41** 1016–1026.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions. *Ann. Math. Statist.* **40** 1950–1979.

ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. Macmillan, New York.

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SHORACK, G. R. (1969). Asymptotic normality of linear combinations of functions of order statistics. *Ann. Math. Statist.* **40** 2041–2050.

SHORACK, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.
SIDDIQUI, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *J. Res. Nat. Bur. Standards* **64B** 145–150.
STIGLER, S. M. (1969). Linear functions of order statistics. *Ann. Math. Statist.* **40** 770–788.
STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477.
STIGLER, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2** 676–693. Correction, **7** 466 (1979).
WELSH, A. H. (1987). The trimmed mean in the linear model (with discussion). *Ann. Statist.* **15** 20–45.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637